International Journal of Population Data Science

Journal Website: www.ijpds.org

Blood-Based Biomarkers and Stroke Risk in Ageing Populations: Opportunities and Challenges Across National Cohort Datasets

Abby L.J. Hensel[1,*], Prabasha Rasaputra[1], Therese Chan[1], Davide Liborio Vetrano[2,3], and Amy T. Hsu[1,4]

Submission History

Submitted:	05/03/2026
Accepted:	11/10/2026
Published:	25/11/2026

[1] Bruyère Health Research Institute, Ottawa, Ontario, Canada
[2] Department of Neurobiology, Aging Research Center, Care Sciences and Society, Karolinska Institutet and Stockholm University, Solna, Sweden
[3] Stockholm Gerontology Research Center, Stockholm, Sweden
[4] Department of Family Medicine, University of Ottawa, Ottawa, Ontario, Canada

Abstract

Cerebrovascular diseases, such as stroke, are among the leading causes of morbidity and mortality worldwide, with markedly higher risk observed in older adults. Blood-based biomarkers offer promising tools for early detection and risk stratification, capturing metabolic, inflammatory, and vascular processes that underlie cerebrovascular disease. However, translating biomarker research from laboratory settings to population-level inference depends critically on the data infrastructure available. In this commentary, we review the landscape of national longitudinal cohort studies and health administrative datasets that can support the study of biomarker-stroke associations in ageing populations, discuss their key strengths and limitations, and argue that no single dataset is universally adequate for this purpose. Using the Canadian Longitudinal Study on Aging (CLSA) and the UK Biobank as illustrative case examples, we demonstrate how cohort design, ascertainment methods, and population characteristics shape the biomarker associations observed. We conclude by identifying opportunities for future work, including data linkage, federated analyses, and the integration of stroke-specific registries, to advance equitable and population-relevant biomarker-stroke research.

Keywords
ageing, biomarkers, CLSA, cohort studies, national health datasets, stroke, UK Biobank

*Corresponding Author:
Email Address: ahensel@bruyere.org (Abby Hensel)

https://doi.org/10.23889/ijpds.v11i1.3445
November 2026 © The Authors. Open Access under CC BY 4.0 (https://creativecommons.org/licenses/by/4.0/deed.en)

Introduction

Cerebrovascular diseases, such as stroke, remain one of the leading causes of mortality and long-term disability worldwide, with approximately 142 incident events per 100,000 individuals globally each year [1]. Stroke risk doubles with each decade after age 55 [2], largely reflecting the accumulation of vascular and metabolic risk factors such as hypertension, diabetes, and heart diseases [3]. As older adult populations continue to grow, the burden of stroke and its associated healthcare and social costs is expected to rise [4, 5], and prevention increasingly depends on the early detection of stroke risk factors and improved risk stratification.

Blood-based biomarkers represent an attractive tool for this purpose. Markers such as glycated haemoglobin (HbA1c) and lipid fractions have long been used in cardiovascular risk assessment [6–8], while inflammatory markers including interleukin-6 (IL-6), high-sensitivity C-reactive protein (hs-CRP), and tumour necrosis factor (TNF) capture systemic processes relevant to vascular and neurodegenerative diseases [9–11]. Small but measurable changes in these biomarkers can signal subclinical disease and refine existing risk prediction models [12–16].

Translating biomarker science to population health, however, requires data infrastructure that combines biological samples with rich phenotypic and longitudinal data. National cohort studies and health administrative datasets have emerged as critical resources for this purpose, but they vary substantially in design, representativeness, biomarker availability, and outcome ascertainment, and not all are equally suited to biomarker-stroke research.

In this commentary, we review the landscape of national datasets relevant to biomarker-stroke research in ageing populations, highlight key strengths and limitations of each, and argue that researchers must carefully consider which resource best fits their specific question. Using the Canadian Longitudinal Study on Aging (CLSA) and the UK Biobank as illustrative case examples, we demonstrate how dataset features shape observed associations. Our goal is not to present definitive causal findings, but to contextualise the choices researchers face and identify opportunities for the field to advance.

The value of biomarkers in population health research

Blood-based biomarkers provide a powerful and minimally invasive window into biological processes relevant to stroke [17–21]. Population-level biomarker research gains statistical power by leveraging large cohorts, which enables the detection of less common exposures and outcomes and the illumination of risk patterns across population subgroups. From a stroke risk perspective, key markers studied at the population-level include metabolic markers (HbA1c, lipids, triglycerides), inflammatory markers (hs-CRP, IL-6, TNF), and emerging markers such as lipoprotein(a) and remnant cholesterol [16]. However, an important conceptual distinction exists between biomarkers of lifetime stroke risk, such as hs-CRP and HbA1c, which reflect long-term vascular burden and are suited to prospective epidemiological studies, and markers of acute stroke, such as glial fibrillary acidic protein (GFAP) and S100 calcium-binding protein B (S100B), which reflect acute neurological injury and are better suited to clinical diagnostic contexts [22]. These markers have been incorporated into risk tools such as the age, biomarkers, and clinical history stroke risk score for atrial fibrillation (ABC-AF score) [23], but their systematic integration into population-level surveillance remains limited. Population health datasets are primarily positioned to investigate the former, and the biomarkers discussed in this commentary reflect that focus.

Population-based datasets for biomarker-stroke research

Longitudinal cohort studies

National longitudinal cohort studies are among the most powerful resources for studying chronic disease aetiology, combining prospective design, repeated data collection, and multidimensional phenotypic data with biospecimen collection [24, 25]. The CLSA [26, 27] and the UK Biobank [28, 29] exemplify this type of resource. Similar potentially relevant cohorts exist across the globe, including the Health and Retirement Study (HRS) in the United States [30], the Mexican Health and Aging Study (MHAS) [31], the China Health and Retirement Longitudinal Study (CHARLS) [32], the Longitudinal Ageing Study in India (LASI) [33], and the Survey of Health, Ageing and Retirement in Europe (SHARE), which spans more than 25 European countries [34]. General population cohorts, such as the Scottish Longitudinal Study (SLS) [35], the Whitehall II study [36], and the Stockholm part of the National study on Aging and Care (SNAC-K) offer complementary perspectives through administrative record linkage and well-characterised socioeconomic gradients in cardiovascular disease. This list is not exhaustive, and the extent to which individual cohorts have been used to study blood-based biomarker-stroke associations specifically varies; with the exception of CHARLS, where a growing body of literature has examined biomarker exposures and incident stroke [37–40], such studies remain limited in most other nationally representative longitudinal ageing cohorts, particularly outside of East Asia. Regardless of which cohort is used, researchers should carefully match available biomarker panels to their exposures of interest and consider how cohort-specific characteristics may shape findings.

Health surveys, administrative data, and clinical registries

National health surveys, such as the National Health Interview Survey (NHIS), the National Health and Nutrition Examination Survey (NHANES), and the Scottish Health Survey (SHeS), differ importantly from cohort studies. The cross-sectional or intermittent designs of these surveys limit causal inference, their biospecimen data are often less comprehensive, and follow-up for incident outcomes is generally not embedded. Health administrative databases such as the Clinical Practice Research Datalink (CPRD) in the UK [41] or ICES (formerly the Institute for Clinical Evaluative Sciences) in Ontario [42] offer a complementary perspective

by linking primary care records with hospital admissions and disease registries for millions of patients, providing near-universal healthcare coverage and clinically diverse samples. On the other hand, population-level stroke registries achieve near-complete ascertainment of incident stroke cases with detailed clinical data. These include national-level registries such as the Sentinel Stroke National Audit Programme (SSNAP) in England, Wales, and Northern Ireland, Riksstroke in Sweden [43], and the Canadian Stroke Registry, as well as subnational registries such as the Ontario Stroke Registry [44], which has been linked to other health surveys and administrative databases to study stroke risk. However, these registries lack pre-stroke biomarker exposures, which restricts their prospective value for risk stratification. Linking these datasets to construct a longitudinal cohort offers a complementary approach and represents a promising strategy to address this limitation.

Limitations and challenges of using national datasets for biomarker-stroke research

Participation bias and generalizability

A fundamental challenge for cohort-based biomarker research is participation bias [45]. Individuals who volunteer for large longitudinal cohort studies, including the CLSA and the UK Biobank, tend to be healthier, more educated, and wealthier than the general population [46, 47]. Both cohorts comprise individuals living primarily in urban areas, resulting in the under-representation of populations in rural or remote communities [48]. Minority and marginalised groups may be further underrepresented due to mistrust, language barriers, or logistical challenges [46, 47]. These structural biases mean that biomarker and stroke distributions in these cohorts may not reflect the true population values, and estimates may be distorted if factors influencing participation are also associated with disease risk. Survey weights can partially correct for unequal sampling but cannot fully eliminate bias arising from limited recruitment and retention of underrepresented groups, which reflects barriers in study design and implementation rather than characteristics of these groups. The operational costs of longitudinal cohort studies are also considerable, and the long-term sustainability of cohort infrastructure is not guaranteed.

Outcome ascertainment

The method of stroke outcome ascertainment varies considerably across datasets. Hospital admissions data – as used in the UK Biobank through International Classification of Diseases (ICD)-coded hospital episodes – provide objective, clinically validated diagnoses of high quality [49], but capture only individuals who present to hospital. Mild strokes or transient ischaemic attacks (TIAs) that do not result in admission may be systematically missed, and differential healthcare access can introduce bias [50]. Self-reported health outcomes, as used in the CLSA, avoid some access-related biases but introduce others: previous studies have found substantial discrepancies between self-reported diagnoses and administrative records [51], and self-report depends on participants' memory, health literacy, and prior access to care. Combining self-report with record linkage offers a pragmatic solution, though linkage completeness and quality must be carefully evaluated.

Confounding and cross-cohort comparability

Observed biomarker-stroke associations are also subject to confounding by socioeconomic, behavioural, and healthcare system factors. Socioeconomic status and healthcare access can strongly influence both biomarker profiles and disease outcomes, meaning observed associations may partly reflect systemic inequalities rather than direct biological mechanisms. Cross-cohort comparisons are further complicated by differences in healthcare systems. More broadly, the potential to synthesise findings across cohorts through federated or pooled analyses can be hindered by heterogeneity in measurement protocols, covariate definitions, and outcome ascertainment. Harmonisation efforts, such as those undertaken by the IALSA (Integrative Analysis of Longitudinal Studies of Aging) network, are essential for realising the comparative potential of national cohort infrastructure [52].

Illustrative case examples: the CLSA and the UK biobank

Case example 1: biomarkers and stroke in the CLSA

To illustrate opportunities and limitations described above, we provide a proof-of-concept analysis using the CLSA Comprehensive Cohort data (methods in Appendix A), examining whether blood-based biomarkers measured at baseline were associated with incident stroke or TIA ascertained approximately six years later. The goal was not to establish causal relationships, but to demonstrate how a national longitudinal cohort with rich biomarker data can be used to investigate these associations, and to highlight where dataset characteristics shape the findings.

Consistent with established stroke epidemiology, participants reporting stroke/TIA were older and exhibited greater socioeconomic disadvantage and comorbidity (Appendix B). In unadjusted comparisons (Table 1), median HbA1c, IL-6, and hs-CRP were higher and median total cholesterol and low-density lipoprotein (LDL) cholesterol were lower among individuals with stroke/TIA than among those without. Differences in high-density lipoprotein (HDL) cholesterol, triglycerides, and TNF were not statistically significant. After adjustment, HbA1c, triglycerides, and TNF remained independently associated with stroke/TIA, suggesting these markers capture risk dimensions beyond traditional covariates (Figure 1), while associations with IL-6, total cholesterol, and LDL cholesterol were attenuated and no longer statistically significant. In both unadjusted and adjusted regression models, hs-CRP was not significantly associated with stroke/TIA (Appendix Table B3). Sex-stratified analyses showed that the adjusted associations of HbA1c and TNF persisted among men but not women (Appendix Table

Table 1: Biomarker values by incident stroke/TIA status in the CLSA; weighted

Characteristic	Overall, N=26,156	Stroke/TIA, N=822	No stroke/TIA, N=25,334	p-value[1]
HbA1c[2]	5.50 (5.20, 5.70)	5.60 (5.30, 5.90)	5.40 (5.20, 5.70)	**<0.001**
Missing	2,600	96	2,504	
Total cholesterol (mmol/L)	5.19 (4.48, 5.88)	5.07 (4.21, 5.68)	5.19 (4.48, 5.88)	**0.015**
Missing	2,500	93	2,407	
High-density lipoprotein (mmol/L)	1.41 (1.13, 1.75)	1.34 (1.10, 1.69)	1.41 (1.14, 1.75)	0.200
Missing	2,500	93	2,407	
Low-density lipoprotein (mmol/L)	2.88 (2.27, 3.52)	2.71 (2.07, 3.28)	2.88 (2.27, 3.52)	**0.002**
Missing	3,015	113	2,902	
Triglycerides (mmol/L)	1.55 (1.09, 2.23)	1.62 (1.24, 2.33)	1.55 (1.09, 2.23)	0.110
Missing	2,500	93	2,407	
Interleukin-6 (pg/mL)	1.83 (1.17, 2.93)	2.78 (1.64, 4.30)	1.81 (1.15, 2.87)	**<0.001**
Missing	17,479	547	16,932	
TNF (pg/mL)	0.97 (0.78, 1.20)	1.01 (0.78, 1.33)	0.96 (0.78, 1.19)	0.200
Missing	17,782	558	17,224	
hs-CRP (mg/L)	1.20 (0.70, 2.80)	1.50 (0.90, 3.00)	1.20 (0.70, 2.70)	**0.009**
Missing	2,500	93	2,407	

[1] p-values presented to 3 decimal places. Wilcoxon rank-sum test for complex survey samples. Bold values indicate $p<0.05$.
[2] HbA1c: glycated haemoglobin, hs-CRP: high-sensitivity C-reactive protein, TIA: transient ischaemic attack, TNF: tumour necrosis factor.
Values are weighted median (interquartile range). Biomarker values reflect measurements taken at CLSA baseline (2011-2015), prior to stroke/TIA ascertainment at second follow-up (2018-2021).

B4), whereas the adjusted estimate for triglycerides was identical in men and women (OR: 1.11; 95% CI: 0.99-1.24 and 0.96-1.28, respectively), with a wider interval among women. Differences in statistical significance between sexes should be interpreted with caution given the smaller number of stroke/TIA events among women in the CLSA and the potential for differential treatment patterns and comorbidity profiles. The self-reported outcome ascertainment in the CLSA, combined with the predominantly white and urban composition of the Comprehensive Cohort, limits the generalizability of these findings even after applying CLSA sampling weights, which account for the sampling design and non-response and calibrate the sample to the Canadian population (Appendix A). Compared to another Canadian population-based study on stroke incidence, for example, the CLSA cohort included a larger proportion of white and urban-residing individuals [3, 53]. These results are therefore best understood as illustrative of what can and cannot be learned from this resource, rather than as definitive estimates of biomarker-stroke associations in the older Canadian population?

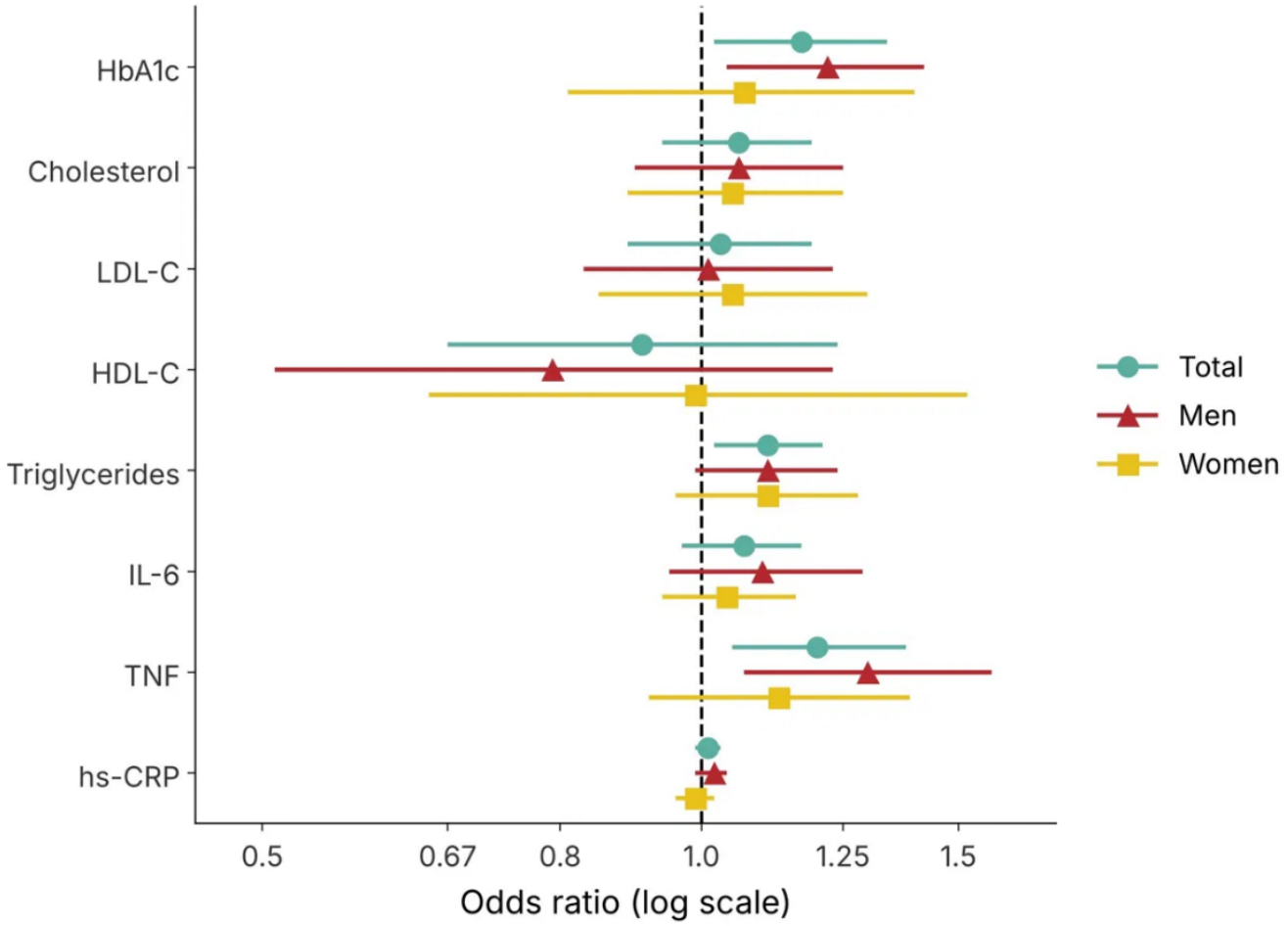
Case example 2: cross-cohort comparison with the UK biobank

Comparing the CLSA findings with UK Biobank studies highlight how dataset features shape results. Analyses using data from the UK Biobank found a strong, persistent association between elevated hs-CRP and incident ischaemic stroke over more than a decade of follow-up [54] (Appendix C, Table C1), in contrast to the CLSA where this association attenuated substantially with adjustment, likely reflecting UK Biobank's larger sample and objective hospital-linked outcome ascertainment. Sex-stratified UK Biobank analyses reported stronger relative associations for traditional risk factors among women despite higher absolute stroke rates in men [55] (Appendix C Figures C1-C3), while the CLSA associations of HbA1c and TNF persisted only among men. This divergence likely reflects the CLSA's older baseline population (mean age ∼63 at baseline vs. ∼56 in the UK Biobank), fewer female stroke/TIA events, and reliance on self-reported outcomes. These are not simply methodological inconveniences but substantively important features: dataset characteristics introduce systematic differences in what can be detected and in whom.

Opportunities for future work

Despite these limitations, the landscape for biomarker-stroke research using national datasets offers genuine opportunities. First, linkage between population cohort data and clinical registries such as SSNAP or Riksstroke could substantially improve outcome ascertainment while retaining the rich pre-stroke exposure data, enabling researchers to examine biomarker profiles alongside stroke subtype, severity, and functional outcome. The Riksstroke registry, with its existing linkage to Swedish national registers, provides a working

Figure 1: Total sample and sex-stratified binomial logistic regression models demonstrating adjusted associations between blood biomarkers and 6-year incident stroke/TIA in the CLSA; weighted. Results are presented as adjusted odds ratios (ORs) with 95% confidence intervals plotted on a logarithmic scale. Each biomarker was modelled as a continuous variable (per unit increase). Adjusted models control for age, race/ethnicity, heart disease, diabetes, alcohol consumption, smoking status, physical activity, body mass index, household income, education, and urban/rural residence; sex was additionally adjusted for in the total sample model. Biomarker values reflect measurements taken at CLSA baseline (2011-2015), prior to stroke/TIA ascertainment at second follow-up (2018-2021). Stroke and TIA outcomes were combined due to limited event counts



model for this approach [43]. Second, federated analyses involving data from multiple national cohorts offer a path forward for examining disease aetiology within a meta-analytic framework. While individual cohorts may have limitations, the commonalities that exist across many national longitudinal cohort studies provide an opportunity to increase statistical power for examining the associations of interest while resolving discrepancies across individual studies. Where possible, harmonisation of biomarker measurement protocols and outcome definitions across cohorts will be essential for this effort. Third, future studies should prioritise recruitment and retention of underrepresented populations – including non-white ethnic groups, rural residents, and individuals with lower socioeconomic status – through oversampling strategies, community-based recruitment, and culturally appropriate participation support

Finally, the integration of genomic, proteomic, and metabolomic data with existing cohort infrastructure opens new possibilities for understanding the biological mechanisms linking systemic inflammation and metabolic dysregulation to stroke risk. Real-world tools such as the Stroke Riskometer (https://nisan.aut.ac.nz/stroke-riskometer) demonstrate that personalised, multi-input risk assessment is feasible at scale. Incorporating richer biomarker panels into such tools, validated across diverse national cohorts, represents an important next step for translating population-based biomarker research into actionable prevention strategies.

Conclusions

No single national dataset is universally adequate for studying the association between blood-based biomarkers and stroke risk in ageing populations. The choice of dataset profoundly shapes the findings obtained, through participation bias, outcome ascertainment, demographic composition, and confounding structure. The CLSA and UK Biobank are powerful and complementary resources, but both are subject to important limitations that constrain the populations and associations that can be studied. Other national datasets offer different trade-offs and should be selected based on the specific research question at hand.

Despite these challenges, national longitudinal cohort studies enriched with biomarker measures provide the most suitable currently available platforms for investigating how biological pathways and social determinants jointly contribute to stroke risk across populations. Using CLSA data as a proof-of-concept, we demonstrated that metabolic and inflammatory biomarkers are associated with incident stroke/TIA in a Canadian ageing cohort, with patterns that partly differ from those observed in the UK Biobank. These differences are themselves informative, highlighting how cohort context shapes biomarker-disease relationships, but the similarities across these two data platforms also demonstrate the potential for future research to advance this area through federated data analysis and a meta-analytic approach. Moving forward, efforts to link cohort data to clinical registries, harmonise measurement protocols across datasets, and recruit more representative samples will be essential to advancing equitable, population-relevant biomarker-stroke research and reducing the burden of stroke in ageing populations.

Acknowledgements

This research was made possible using the data/biospecimens collected by the Canadian Longitudinal Study on Aging (CLSA). Funding for the CLSA is provided by the Government of Canada, through the Canadian Institutes of Health Research (CIHR) under grant reference no. LSA 94473, and the Canada Foundation for Innovation, as well as the following provinces: Newfoundland, Nova Scotia, Quebec, Ontario, Manitoba, Alberta, and British Columbia. This research has been conducted using the CLSA Comprehensive Baseline Dataset version 7.0, Comprehensive Follow-up 1 Dataset version 5.0, and Comprehensive Follow-up 2 Dataset version 2.0 under Application ID 2310018. The CLSA is led by Drs. Parminder Raina, Christina Wolfson and Susan Kirkland. The time and commitment of the participants to the CLSA study platform is gratefully acknowledged, without whom this research would not be possible.

AI tool disclosure

The authors declare that no generative AI tools were used in the preparation of this manuscript.

Conflicts of interests

We have no conflicts of interest to disclose. The opinions expressed in this manuscript are the authors' own and do not reflect the views of the Canadian Longitudinal Study on Aging.

Data availability statement

Data are available from the Canadian Longitudinal Study on Aging (www.clsa-elcv.ca) for researchers who meet the criteria for access to de-identified CLSA data.

Ethics statement

This study was approved by the Bruyère Research Ethics Board (BREB) at Bruyère Health Research Institute, Ottawa, Ontario, Canada (Project ID: M16-24-010).

Funding

This project was funded by a Canadian Institutes of Health Research (CIHR) Catalyst Grant (funding reference no. ACD–200172).

References

1. Feigin VL, Abate MD, Abate YH, ElHafeez SA, Abd-Allah F, Abdelalim A, et al. Global, regional, and national burden of stroke and its risk factors, 1990–2021: A systematic analysis for the Global Burden of Disease Study 2021. Lancet Neurol 2024;23:973–1003. https://doi.org/10.1016/S1474-4422(24)00369-7

2. Yousufuddin M, Young N. Aging and ischemic stroke. Aging 2019;11:2542–4. https://doi.org/10.18632/aging.101931

3. Joundi RA, Patten SB, Williams JV, Smith EE. Vascular risk factors and stroke risk across the life span: A population-representative study of half a million people. Int J Stroke 2022;17:1021–9. https://doi.org/10.1177/17474930211070682

4. Yu AYX, Smith EE, Krahn M, Austin PC, Rashid M, Fang J, et al. Association of neighborhood-level material deprivation with health care costs and outcome after stroke. Neurol 2021;97:e1503-1511. https://doi.org/10.1212/WNL.0000000000012676

5. Strilciuc S, Grad DA, Radu C, Chira D, Stan A, Ungureanu M, et al. The economic burden of stroke: A systematic review of cost of illness studies. J Med Life 2021;14:606–19. https://doi.org/10.25122/jml-2021-0361

6. Boye KS, Thieu VT, Lage MJ, Miller H, Paczkowski R. The association between sustained HbA1c control and long-term complications among individuals with type 2 diabetes: A retrospective study. Adv Ther 2022;39:2208–21. https://doi.org/10.1007/s12325-022-02106-4

7. Arsenault BJ, Boekholdt SM, Kastelein JJP. Lipid parameters for measuring risk of cardiovascular disease. Nat Rev Cardiol 2011;8:197–206. https://doi.org/10.1038/nrcardio.2010.223

8. Sherwani SI, Khan HA, Ekhzaimy A, Masood A, Sakharkar MK. Significance of HbA1c test in diagnosis and prognosis of diabetic patients. Biomark Insights 2016;11:95–104. https://doi.org/10.4137/BMI.S38440

9. Ong KL, Chung RWS, Hui N, Festin K, Lundberg AK, Rye K-A, et al. Usefulness of certain protein biomarkers for prediction of coronary heart disease. Am. J. Cardiol.

2020;125:542–8. https://doi.org/10.1016/j.amjcard.2019.11.016

10. The Emerging Risk Factors Collaboration. C-reactive protein concentration and risk of coronary heart disease, stroke, and mortality: An individual participant meta-analysis. Lancet 2010;375:132–40. https://doi.org/10.1016/S0140-6736(09)61717-7

11. Tuomisto K, Jousilahti P, Sundvall J, Pajunen P, Salomaa V. C-reactive protein, interleukin-6 and tumor necrosis factor alpha as predictors of incident coronary and cardiovascular events and total mortality. A population-based, prospective study. Thromb Haemost 2006;95:511–8. https://doi.org/10.1160/TH05-08-0571

12. Ullah A, Sajid S, Qureshi M, Kamran M, Anwaar MA, Naseem MA, et al. Novel biomarkers and the multiple-marker approach in early detection, prognosis, and risk stratification of cardiac diseases: A narrative review. Cureus 2023;15:e42081. https://doi.org/10.7759/cureus.42081

13. Castelli WP, Abbott RD, McNamara PM. Summary estimates of cholesterol used to predict coronary heart disease. Circ 1983;67:730–4. https://doi.org/10.1161/01.CIR.67.4.730

14. Richardson TG, Sanderson E, Palmer TM, Ala-Korpela M, Ference BA, Smith GD, et al. Evaluating the relationship between circulating lipoprotein lipids and apolipoproteins with risk of coronary heart disease: A multivariable Mendelian randomisation analysis. PLOS Med 2020;17:e1003062. https://doi.org/10.1371/journal.pmed.1003062

15. Kou M, Ding N, Ballew SH, Salameh MJ, Martin SS, Selvin E, et al. Conventional and novel lipid measures and risk of peripheral artery disease. Arterioscler Thromb Vasc Biol 2021;41:1229–38. https://doi.org/10.1161/ATVBAHA.120.315828

16. Wadström BN, Wulff AB, Pedersen KM, Jensen GB, Nordestgaard BG. Elevated remnant cholesterol increases the risk of peripheral artery disease, myocardial infarction, and ischaemic stroke: A cohort-based study. Eur Heart J 2022;43:3258–69. https://doi.org/10.1093/eurheartj/ehab705

17. Dhauria M, Mondal R, Deb S, Shome G, Chowdhury D, Sarkar S, et al. Blood-based biomarkers in Alzheimer's disease: Advancing non-invasive diagnostics and prognostics. Int J Mol Sci 2024;25:10911. https://doi.org/10.3390/ijms252010911

18. Krawczuk D, Kulczyńska-Przybik A, Mroczko B. Clinical application of blood biomarkers in neurodegenerative diseases—Present and future perspectives. Int J Mol Sci 2024;25:8132. https://doi.org/10.3390/ijms25158132

19. Garady L, Soota A, Shouche Y, Chandrachari KP, K V S, Shankar P, et al. A narrative review of the role of blood biomarkers in the risk prediction of cardiovascular diseases. Cureus 2024;16:e74899. https://doi.org/10.7759/cureus.74899

20. Catanese L, Rupprecht H, Huber TB, Lindenmeyer MT, Hengel FE, Amann K, et al. Non-invasive biomarkers for diagnosis, risk prediction, and therapy guidance of glomerular kidney diseases: A comprehensive review. Int J Mol Sci 2024;25:3519. https://doi.org/10.3390/ijms25063519

21. Yang M, Zhang A, Chen M, Cao J. Advances in circulating biomarkers for neurodegenerative diseases, traumatic brain injuries, and central nervous system tumors. Ann Lab Med 2025;45:381–90. https://doi.org/10.3343/alm.2024.0611

22. Jickling GC, Liu D, Stamova B, Ander BP, Zhan X, Lu A, et al. Hemorrhagic transformation after ischemic stroke in animals and humans. J Cereb Blood Flow Metab 2014;34:185–99. https://doi.org/10.1038/jcbfm.2013.203

23. Hijazi Z, Lindbäck J, Alexander JH, Hanna M, Held C, Hylek EM, et al. The ABC (age, biomarkers, clinical history) stroke risk score: A biomarker-based risk score for predicting stroke in atrial fibrillation. Eur Heart J 2016;37:1582–90. https://doi.org/10.1093/eurheartj/ehw054

24. Boehme AK, Esenwa C, Elkind MSV. Stroke risk factors, genetics, and prevention. Circ Res 2017;120:472–95. https://doi.org/10.1161/CIRCRESAHA.116.308398

25. Buka SL, Rosenthal SR, Lacy ME. Epidemiological study designs: Traditional and novel approaches to advance life course health development research. In: Halfon N, Forrest CB, Lerner RM, Faustman EM, editors. Handbook of Life Course Health Development. Cham: Springer International Publishing; 2018. p. 541–60. https://doi.org/10.1007/978-3-319-47143-3_22

26. Raina PS, Wolfson C, Kirkland SA, Griffith LE, Oremus M, Patterson C, et al. The Canadian Longitudinal Study on Aging (CLSA). Canadian Journal on Aging 2009;28:221–9. https://doi.org/10.1017/S0714980809990055

27. Raina P, Wolfson C, Kirkland S, Griffith LE, Balion C, Cossette B, et al. Cohort Profile: The Canadian Longitudinal Study on Aging (CLSA). International Journal of Epidemiology 2019 Dec.;48:1752–1753j. https://doi.org/10.1093/ije/dyz173

28. Collins R. What makes UK Biobank special? Lancet 2012;379:1173–4. https://doi.org/10.1016/S0140-6736(12)60404-8

29. Hewitt J, Walters M, Padmanabhan S, Dawson J. Cohort profile of the UK Biobank: Diagnosis and characteristics of cerebrovascular disease. BMJ Open 2016;6:e009161. https://doi.org/10.1136/bmjopen-2015-009161

30. Sonnega A, Faul JD, Ofstedal MB, Langa KM, Phillips JW, Weir DR. Cohort profile: The Health and Retirement Study (HRS). Int J Epidemiol 2014;43:576–85. https://doi.org/10.1093/ije/dyu067

31. Wong R, Michaels-Obregon A, Palloni A. Cohort profile: The Mexican Health and Aging Study (MHAS). Int J Epidemiol 2017;46:e2. https://doi.org/10.1093/ije/dyu263

32. Zhao Y, Hu Y, Smith JP, Strauss J, Yang G. Cohort profile: The China Health and Retirement Longitudinal Study (CHARLS). Int J Epidemiol 2014;43:61–8. https://doi.org/10.1093/ije/dys203

33. Perianayagam A, Bloom D, Lee J, Parasuraman S, Sekher TV, Mohanty SK, et al. Cohort profile: The Longitudinal Ageing Study in India (LASI). Int J Epidemiol 2022;51:e167–76. https://doi.org/10.1093/ije/dyab266

34. Börsch-Supan A, Brandt M, Hunkler C, Kneip T, Korbmacher J, Malter F, et al. Data resource profile: The Survey of Health, Ageing and Retirement in Europe (SHARE). Int J Epidemiol 2013;42:992–1001. https://doi.org/10.1093/ije/dyt088

35. Boyle PJ, Feijten P, Feng Z, Hattersley L, Huang Z, Nolan J, et al. Cohort profile: The Scottish Longitudinal Study (SLS). Int J Epidemiol 2009;38:385–92. https://doi.org/10.1093/ije/dyn087

36. Marmot M, Brunner E. Cohort profile: The Whitehall II study. Int J Epidemiol 2005;34:251–6. https://doi.org/10.1093/ije/dyh372

37. Wang Q, Zhao L, Zuo T, Peng W. Nonlinear relationship between incidence of new-onset stroke and plasma atherosclerotic index in middle-aged and older adults. Front. Neurol. 2025;16:1597616. https://doi.org/10.3389/fneur.2025.1597616

38. Feng Q, Ren L, Chen S, Li X. Platelet-to-high-density lipoprotein cholesterol ratio as a predictor of stroke risk: A longitudinal analysis of the China health and retirement longitudinal study. Front. Neurol. 2025;16:1503743. https://doi.org/10.3389/fneur.2025.1503743

39. Qi Y, Shang X, Han T, Han N, Jiang Z, Yan H, et al. Serum cystatin C and stroke risk: A national cohort and Mendelian randomization study. Front. Endocrinol. 2024;15:1355948. https://doi.org/10.3389/fendo.2024.1355948

40. Zhang X, Yang L, Jin Z, Li W, Xiang H, Wang W, et al. Frailty index is positively associated with stroke risk in nationally representative cohorts from the united States and China. Sci Rep 2025;15:28440. https://doi.org/10.1038/s41598-025-14116-7

41. Herrett E, Gallagher AM, Bhaskaran K, Forbes H, Mathur R, van Staa T, et al. Data resource profile: Clinical Practice Research Datalink (CPRD). Int J Epidemiol 2015;44:827–36. https://doi.org/10.1093/ije/dyv098

42. Schull MJ, Azimaee M, Marra M, Cartagena RG, Vermeulen MJ, Ho MM, et al. ICES: Data, discovery, better health. Int J Popul Data Sci 2019;4. https://doi.org/10.23889/ijpds.v4i2.1135

43. Appelros P, Jonsson F, Åsberg S, Asplund K, Glader E-L, Åsberg KH, et al. Trends in stroke treatment and outcome between 1995 and 2010: Observations from Riks-Stroke, the Swedish Stroke Register. Cerebrovasc Dis 2013;37:22–9. https://doi.org/10.1159/000356346

44. Yu AYX, Silver FL, Fang J, Hill MD, Lindsay P, Kapral MK. Creation and evolution of the Ontario Stroke Registry: Protocol and two decades of data from a population-based clinical stroke registry. Can. J. Neurol. Sci. 2025;52:920–5. https://doi.org/10.1017/cjn.2025.13

45. Iwasaki M. Commentary: Factors associated with non-participation in cohort studies emphasize the need to generalize the results with care. J Epidemiol 2015;25:89–90. https://doi.org/10.2188/jea.JE20140269

46. Pardhan S, Sehmbi T, Wijewickrama R, Onumajuru H, Piyasena MP. Barriers and facilitators for engaging underrepresented ethnic minority populations in healthcare research: An umbrella review. Int J Equity Health 2025;24:70. https://doi.org/10.1186/s12939-025-02431-4

47. National Academies of Sciences, Engineering, and Medicine; Policy and Global Affairs; Committee on Women in Science, Engineering, and Medicine; Committee on Improving the Representation of Women and Underrepresented Minorities in Clinical Trials and Research. Improving Representation in Clinical Trials and Research: Building Research Equity for Women and Underrepresented Groups. Washington (DC): National Academies Press (US); 2022.

48. Burgard SA, Chen PV. Challenges of health measurement in studies of health disparities. Soc Sci Med 2014;106:143–50. https://doi.org/10.1016/j.socscimed.2014.01.045

49. Boulanger V, MacLaurin A, Quach C. Barriers and facilitators for using administrative data for surveillance purpose: A narrative overview. J Hosp Infect 2025;155:25–36. https://doi.org/10.1016/j.jhin.2024.09.027

50. Brennan P, Perola M, van Ommen G-J, Riboli E, On behalf of the European Cohort Consortium. Chronic disease research in Europe and the need for integrated population cohorts. Eur J Epidemiol 2017;32:741–9. https://doi.org/10.1007/s10654-017-0315-2

51. Muggah E, Graves E, Bennett C, Manuel DG. Ascertainment of chronic diseases using population health data: A comparison of health administrative data and patient self-report. BMC Public Health 2013;13:16. https://doi.org/10.1186/1471-2458-13-16

52. Hofer SM, Piccinin AM. Integrative data analysis through coordination of measurement and analysis protocol across independent longitudinal studies. Psychol Methods 2009;14:150–64. https://doi.org/10.1037/a0015566

53. Kapral MK, Austin PC, Jeyakumar G, Hall R, Chu A, Khan AM, et al. Rural-urban differences in stroke risk factors, incidence, and mortality in people with and without prior stroke. Circ Cardiovasc Qual Outcomes 2019;12:e004973. https://doi.org/10.1161/CIRCOUTCOMES.118.004973

54. Wang Y, Zhang X, Liang A, Niu Y, Chen D, Du Z, et al. High-sensitivity C-reactive protein and risk of retinal artery occlusion and ischaemic stroke: A cross-cohort study. Br J Ophthalmol 2025;109:1081–7. https://doi.org/10.1136/bjo-2023-325044

55. Peters SAE, Carcel C, Millett ERC, Woodward M. Sex differences in the association between major risk factors and the risk of stroke in the UK Biobank cohort study. Neurol 2020;95:e2715-2726. https://doi.org/10.1212/WNL.0000000000010982

Abbreviations

HbA1c:	glycated haemoglobin
HDL-C:	high-density lipoprotein cholesterol
hs-CRP:	high-sensitivity C-reactive protein
LDL-C:	low-density lipoprotein cholesterol
TIA:	transient ischaemic attack
TNF:	tumour necrosis factor

Appendix A: CLSA analysis methods

Data source and sample overview

We used data from baseline (2011-2015), first follow-up (2015-2018), and second follow-up (2018-2021) of the Canadian Longitudinal Study on Aging (CLSA) Comprehensive cohort. CLSA is an ongoing nationwide cohort study, with 51,338 participants from 10 Canadian provinces aged 45-85 years old at the time of recruitment (2011-2015). Participants provided data on demographic, biological, medical, psychosocial, economic, lifestyle, and behavioural factors through either a computer-assisted telephone survey or an in-person interview. Of the 51,338 participants, a subset of 30,097 (Comprehensive cohort) living within 25 to 50 km of one of eleven data collection sites located in seven provinces were invited to complete detailed physical assessments. From this cohort, we assessed a sample of 26,093 participants who had blood-biomarker data at baseline and stroke or transient ischaemic attack (TIA) data at the second follow-up. A detailed description of CLSA recruitment and data collection procedures has been previously described elsewhere [1, 2].

Blood-based biomarkers

We used data on eight blood biomarkers assessed at the CLSA baseline visit (2011-2015), including HbA1c, total cholesterol, high-density lipoprotein (HDL-C), low-density lipoprotein (LDL-C), triglycerides, interleukin-6 (IL-6), tumour necrosis factor (TNF), and high-sensitivity C-reactive protein (hs-CRP). All biomarker values used in analyses therefore reflect exposure status at baseline, prior to stroke/TIA ascertainment at follow-up 2.

Biological specimens, including blood samples, were collected at participants' visits to a CLSA data collection site by trained phlebotomists following standardised protocols. Blood was collected into ethylenediaminetetraacetic acid (EDTA) tubes for plasma and serum separator tubes as appropriate to each assay. Plasma and serum were separated by centrifugation and stored at $-80^{\circ}C$ until analysis. HbA1c was measured using high-performance liquid chromatography (HPLC); lipid fractions (total cholesterol, HDL-C, LDL-C, triglycerides) were measured by enzymatic colorimetric methods; IL-6 and TNF were measured by electrochemiluminescence immunoassay; and hs-CRP was measured by high-sensitivity immunoturbidimetric assay. Further details of the CLSA biospecimen collection and laboratory procedures are described elsewhere [1].

Incident stroke/TIA

Participants reporting a physician-diagnosed stroke or TIA at baseline were excluded from the analytical sample. Incident stroke/TIA was ascertained at the first and second follow-up using self-report. Stroke was defined by the question: "Has a doctor ever told you that you have experienced a stroke or cerebrovascular accident? (CVA)?" Likewise, TIA was defined by the question: "Has a doctor ever told you that you have experienced a mini-stroke or TIA?" Participants who responded yes to either question at follow-up 1 or follow-up 2 (but not at baseline) were classified as incident stroke/TIA cases. Stroke and TIA were combined into a single outcome due to the relatively small number of events in the analytical sample, which precluded reliable subtype-specific analyses.

Covariates

Sociodemographic and clinical covariates were selected based on established associations with stroke risk in the literature. Selected covariates included age, biological sex, educational level, annual household income, race/ethnicity, residence (i.e., urban vs. rural), body mass index (BMI), heart disease, and diabetes. Due to the small representation of individual non-white racial/ethnic groups in the CLSA cohort ($<6\%$), race/ethnicity was categorised as white versus non-white to ensure adequate cell sizes and robust statistical analyses.

Statistical analysis

Descriptive statistics were used to summarise characteristics of the study samples stratified by stroke/TIA status. Proportions and means with standard deviations (SDs) using sampling (inflation) weights were used to summarise continuous and categorical variables, respectively. Baseline participant characteristics and biomarker concentrations were compared between incident stroke/TIA cases and non-cases using Wilcoxon rank-sum test for complex survey samples (continuous variables) and the chi-squared test with Rao & Scott's second-order correction (categorical variables). All descriptive analyses incorporated CLSA sampling weights.

Sampling weights provided by the CLSA were used. The CLSA calculates these weights to account for the study's sampling design and for non-response, and calibrates them to national census-based population estimates so that weighted estimates reflect the corresponding Canadian population aged 45-85 years. Two types of weights are available. Inflation weights, which sum to the size of the target population, were applied to the descriptive statistics, whereas analytic weights, which are proportional to the inflation weights but rescaled to the sample size, were applied to the logistic regression models.

Binomial logistic regression models using analytic weights were used to assess the association between each biomarker level and 6-year incident stroke/TIA (binary outcome). A crude (unadjusted) model was first estimated for each biomarker. An adjusted model was then estimated, controlling for age, sex, race/ethnicity, heart disease, stroke, diabetes, alcohol consumption, smoking status, physical activity, BMI, income, education, and urban/rural residence. All biomarkers were modelled as continuous variables, with units as reported in Table 1 and Appendix B. Sex-stratified analyses were conducted by repeating these models separately for males and females. Results are reported as odds ratios (OR) with 95% confidence interval (CI) estimates. Analyses were conducted using R (R Core Team, 2022).

References

1. Raina PS, Wolfson C, Kirkland SA, Griffith LE, Oremus M, Patterson C, et al. The Canadian Longitudinal Study on Aging (CLSA). Canadian Journal on Aging/La

Revue canadienne du vieillissement 2009 Sept.;28:221–9. https://doi.org/10.1017/S0714980809990055

2. Raina P, Wolfson C, Kirkland S, Griffith LE, Balion C, Cossette B, et al. Cohort Profile: The Canadian Longitudinal Study on Aging (CLSA). International Journal of Epidemiology 2019 Dec.;48:1752–1753j. https://doi.org/10.1093/ije/dyz173

Appendix B: CLSA results tables

Table B1: Sociodemographic characteristics of sample cohort (N=26,093); Unweighted count (Weighted prevalence)

Characteristic	Overall, N=26,156	Stroke/TIA, N=822	No stroke/TIA, N=25,334	p-value[1]
Age	61 (54, 69)	70 (62,78)	61 (54, 69)	**<0.001**
Sex				0.013
Female	13,366 (53%)	385 (48%)	12,981 (53%)	
Male	12,790 (47%)	437 (52%)	12,353 (47%)	
Race				0.300
White	25,065 (94%)	794 (93%)	24,271 (94%)	
Non-white	1,091 (5.8%)	28 (6.7%)	1,063 (5.8%)	
Education				**<0.001**
Less than secondary school graduation	1,244 (15%)	67 (27%)	1,177 (15%)	
Secondary school graduation, no post-secondary education	2,350 (11%)	86 (11%)	2,264 (11%)	
Some post-secondary education	1,879 (9.1%)	<80 (<15%)	1,804 (9.0%)	
Post-secondary degree/diploma	20,644 (64%)	592 (51%)	20,052 (65%)	
Missing	39 (0.1%)	<10 (<0.1%)	37 (0.1%)	
Marital status				**<0.001**
Single	2,272 (8.5%)	59 (13%)	2,213 (8.4%)	
Married/Common-law relationship	18,331 (75%)	534 (66%)	17,797 (76%)	
Widowed/Separated/Divorced	<5,550 (16%)	229 (22%)	<5,320 (16%)	
Missing	<10 (<0.1%)	-	<10 (<0.1%)	
Urban/rural classification				0.100
Rural	2,136 (5.2%)	55 (4.8%)	2,080 (5.3%)	
Urban	23,705 (94%)	751 (94%)	22,954 (94%)	
Missing	315 (1.0%)	15 (1.4%)	300 (1.0%)	
Household income				**<0.001**
< $50,000	6,332 (25%)	290 (43%)	6,042 (24%)	
$50,000-$100,000	13,791 (52%)	408 (39%)	13,383 (52%)	
≥$100,000	4,464 (17%)	74 (10%)	4,390 (18%)	
Missing	1,569 (5.9%)	50 (6.9%)	1,519 (5.9%)	
Co-resident status				**<0.001**
Living along	5,602 (15%)	235 (22%)	3,367 (15%)	
Living with 1 co-resident	12,719 (44%)	445 (55%)	12,274 (44%)	
Living with 2+ co-residents	7,820 (40%)	142 (23%)	7,678 (41%)	
Missing	15 (<0.1%)	-	15 (<0.1%)	

HbA1c: glycated haemoglobin, TIA: transient ischaemic attack.
[1]p-values presented to 3 decimal places. Wilcoxon rank-sum test for complex survey samples; chi-squared test with Rao & Scott's second-order correction. Bold values indicate p<0.05.

Table B2: Clinical characteristics of sample cohort (N = 26,093); unweighted count (weighted prevalence)

Characteristic	Overall, N=26,156	Stroke/TIA, N=822	No stroke/TIA, N=25,334	p-value[1]
Chronic heart disease				**<0.001**
Yes	2,653 (9.4%)	184 (25%)	2,469 (8.9%)	
No	23,369 (90%)	624 (73%)	22,745 (91%)	
Missing	134 (0.4%)	14 (1.0%)	120 (0.4%)	
Diabetes				0.1
Yes	4,291 (16%)	184 (20%)	4,107 (16%)	
No	21,780 (84%)	<630 (79%)	21,151 (84%)	
Missing	85 (0.5%)	<10 (0.4%)	76 (0.5%)	
Angina				**<0.001**
Yes	993 (3.4%)	66 (10%)	927 (3.2%)	
No	25,042 (96%)	741 (88%)	24,301 (96%)	
Missing	121 (0.6%)	15 (1.8%)	106 (0.5%)	
Myocardial infarction				**<0.001**
Yes	1,052 (3.5%)	76 (11%)	976 (3.3%)	
No	25,004 (96%)	730 (88%)	24,274 (96%)	
Missing	100 (0.4%)	16 (1.4%)	84 (0.4%)	
Alzheimer's/Dementia				0.5
Yes	28 (<0.1%)	0 (0%)	28 (0.1%)	
No	26,068 (<100%)	<820 (<100%)	25,254 (<100%)	
Missing	60 (0.2%)	<10 (<0.5%)	52 (0.2%)	
Parkinson's Disease				0.5
Yes	92 (0.3%)	<10 (<0.5%)	89 (0.3%)	
No	26,013 (99%)	<820 (<100%)	25,202 (99%)	
Missing	51 (0.2%)	<10 (<0.5%)	43 (0.2%)	
Memory problems				**0.010**
Yes	333 (1.5%)	26 (3.5%)	307 (1.4%)	
No	25,751 (98%)	786 (96%)	24,965 (98%)	
Missing	72 (0.3%)	10 (0.6%)	62 (0.3%)	
Mood disorder				**<0.001**
Yes	4,366 (18%)	<180 (<30%)	4,187 (17%)	
No	21,703 (82%)	634 (70%)	21,069 (82%)	
Missing	87 (0.4%)	<10 (<0.7%)	78 (0.4%)	
Clinical depression				**0.003**
Yes	4,179 (16%)	166 (26%)	4,013 (16%)	
No	21,866 (83%)	646 (73%)	21,220 (84%)	
Missing	111 (0.4%)	10 (0.6%)	101 (0.4%)	
Self-rated mental health				**<0.001**
Excellent/Very good/Good	24,843 (95%)	749 (87%)	24,094 (95%)	
Fair/Poor	1,297 (5.3%)	73 (13%)	1,224 (5.1%)	
Missing	16 (<0.1%)	0 (0%)	16 (<0.1%)	
Functional impairment				**<0.001**
No functional impairment	23,931 (92%)	686 (86%)	23,245 (93%)	
Mild/moderate/severe functional impairment	2,142 (6.9%)	132 (13%)	2,010 (6.7%)	
Total functional impairment	10 (<0.1%)	<10 (<0.5%)	<10 (<0.5%)	
Missing	73 (0.6%)	<10 (<0.5%)	71 (0.6%)	

TIA: transient ischaemic attack.

[1] p-values presented to 3 decimal places. Wilcoxon rank-sum test for complex survey samples; chi-squared test with Rao & Scott's second-order correction. Bold values indicate p<0.05.

Table B3: Binomial logistic regression models demonstrating crude and adjusted association between blood biomarkers and incident stroke or transient ischaemic attack; weighted

	Unadjusted OR [95% CI	Adjusted OR [95% CI]
HbA1c (per unit)	1.26 [1.15, 1.37]	1.17 [1.02, 1.34]
Total cholesterol (mmol/L)	0.88 [0.79, 0.99]	1.06 [0.94, 1.19]
Low-density lipoprotein (mmol/L)	0.83 [0.73, 0.95]	1.03 [0.89, 1.19]
High-density lipoprotein (mmol/L)	0.83 [0.63, 1.10]	0.91 [0.67, 1.24]
Triglycerides (mmol/L)	1.10 [1.01, 1.20]	1.11 [1.02, 1.21]
Interleukin-6 (pg/mL)	1.17 [1.09, 1.25]	1.07 [0.97, 1.17]
TNF (pg/mL)	1.21 [1.06, 1.38]	1.20 [1.05, 1.38]
hs-CRP (pg/mL)	1.01 [0.99, 1.03]	1.01 [0.99, 1.03]

CI: confidence interval, HbA1c: glycated haemoglobin, hs-CRP: high-sensitivity C-reactive protein, OR: odds ratio, TNF: tumour necrosis factor.
All biomarkers modelled as continuous variables per unit increase. Adjusted model controls for age, sex, race/ethnicity, heart disease, diabetes, alcohol consumption, smoking status, physical activity, body mass index, household income, education, and urban/rural residence. Rows ordered by biomarker category: metabolic markers (HbA1c, cholesterol fractions, triglycerides) followed by inflammatory markers (interleukin-6, TNF, hs-CRP).

Table B4: Sex-stratified binomial logistic regression models demonstrating crude and adjusted associations between blood biomarkers and incident stroke or transient ischaemic attack; weighted

	Males		Females	
	Unadjusted OR [95% CI]	Adjusted OR [95% CI]	Unadjusted OR [95% CI]	Adjusted OR [95% CI]
HbA1c	1.29 [1.16, 1.44]	1.22 [1.04, 1.42]	1.19 [1.03, 1.39]	1.07 [0.81, 1.40]
Total cholesterol	0.85 [0.73, 0.99]	1.06 [0.90, 1.25]	0.94 [0.80, 1.11]	1.05 [0.89, 1.25]
Low-density lipoprotein	0.76 [0.65, 0.90]	1.01 [0.83, 1.23]	0.92 [0.74, 1.13]	1.05 [0.85, 1.30]
High-density lipoprotein	0.77 [0.52, 1.14]	0.79 [0.51, 1.23]	0.93 [0.62, 1.40]	0.99 [0.65, 1.52]
Triglycerides	1.08 [0.97, 1.19]	1.11 [0.99, 1.24]	1.12 [0.97, 1.28]	1.11 [0.96, 1.28]
Interleukin-6	1.19 [1.08, 1.32]	1.10 [0.95, 1.29]	1.14 [1.04, 1.25]	1.04 [0.94, 1.16]
TNF	1.31 [1.12, 1.53]	1.30 [1.07, 1.58]	1.11 [0.88, 1.39]	1.13 [0.92, 1.39]
hs-CRP	1.02 [1.00, 1.04]	1.02 [0.99, 1.04]	0.99 [0.96, 1.02]	0.99 [0.96, 1.02]

CI: confidence interval, HbA1c: glycated haemoglobin, hs-CRP: high-sensitivity C-reactive protein, OR: odds ratio, TNF: tumour necrosis factor.
All biomarkers modelled as continuous variables per unit increase. Adjusted model controls for age, sex, race/ethnicity, heart disease, diabetes, alcohol consumption, smoking status, physical activity, body mass index, household income, education, and urban/rural residence. Rows ordered by biomarker category: metabolic markers (HbA1c, cholesterol fractions, triglycerides) followed by inflammatory markers (interleukin-6, TNF, hs-CRP).

Appendix C: results from UK biobank studies

Table C1: Multivariable-adjusted hazard ratios (HRs) [95% CI] for the association of hs-CRP and incident ischaemic stroke and transient ischaemic attack

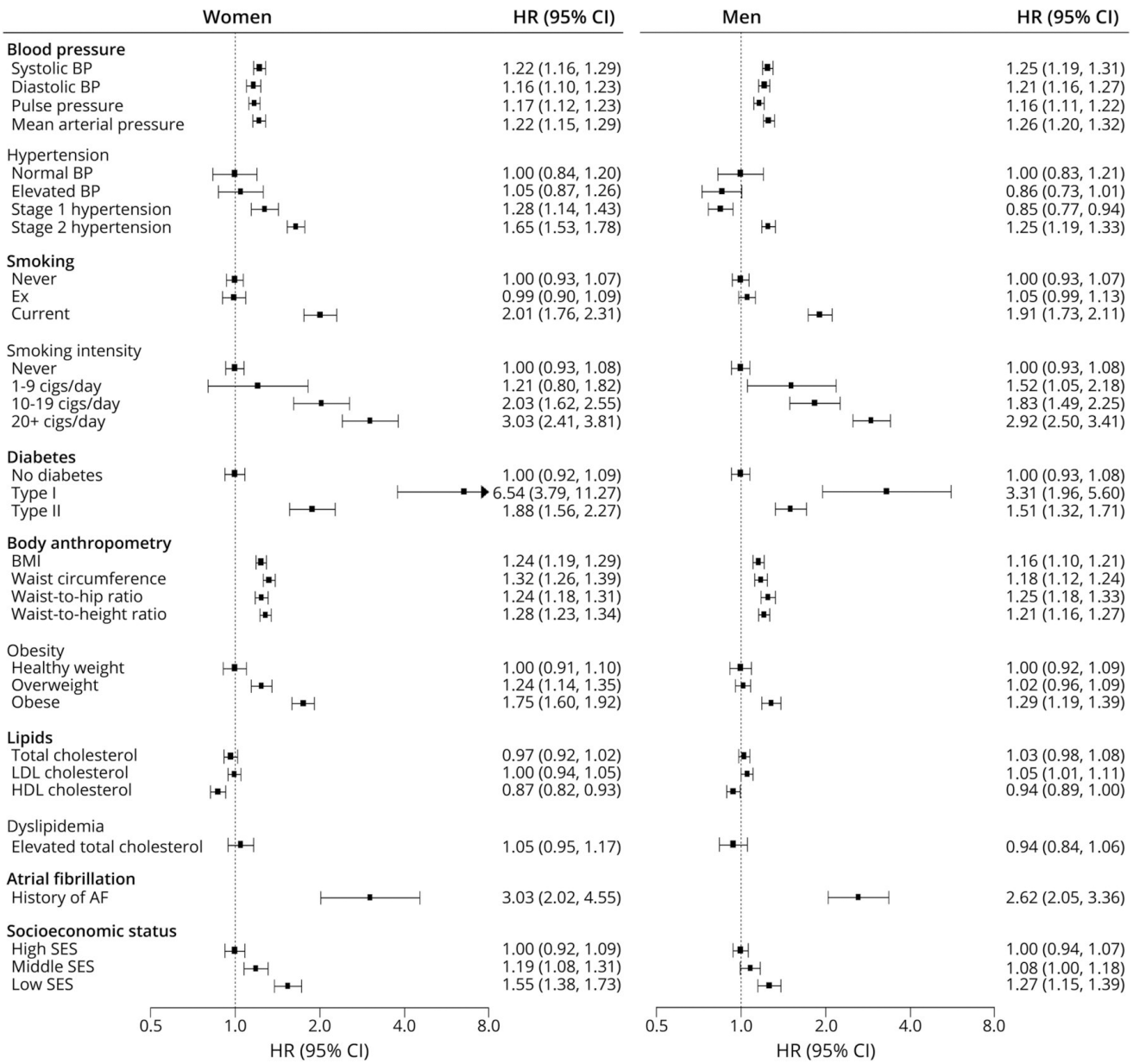
hs-CRP concentration		Incident retinal artery occlusion (n=136)		Incident ischaemic stroke (n=3206)		Incident transient ischaemic attack (n = 1489)	
		Model 1 HR (95% CI)	Model 2 HR (95% CI)	Model 1 HR (95% CI)	Model 2 HR (95% CI)	Model 1 HR (95% CI)	Model 2 HR (95% CI)
Per 10 mg/L		**1.32 (1.03 to 1.69)**	**1.43 (1.01 to 1.76)**	**1.35 (1.29 to 1.42)**	**1.24 (1.17 to 1.33)**	**1.25 (1.15 to 1.36)**	**1.12 (1.01 to 1.25)**
Category	Low	Reference	Reference	Reference	Reference	Reference	Reference
	Median	**1.89 (1.20 to 2.96)**	1.52 (0.91 to 2.52)	**1.41 (1.28 to 1.55)**	1.09 (0.98 to 1.21)	**1.40 (1.22 to 1.60)**	1.15 (0.99 to 1.34)
	High	**1.80 (1.14 to 2.83)**	1.32 (0.77 to 2.29)	**2.01 (1.84 to 2.20)**	**1.48 (1.33 to 1.64)**	**1.81 (1.59 to 2.06)**	**1.40 (1.20 to 1.64)**

Model 1 is the crude model; model 2 is adjusted for age, gender, ethnicity, education, Townsend deprivation index, physical activity, body mass index, cigarette smoking, alcohol drinking, family history of stroke, diabetes mellitus, hypertension and hyperlipidaemia. Bolded items are statistically significant ($p<0.05$).

hs-CRP, high-sensitivity C-reactive protein.

Note: From *High-sensitivity C-reactive protein and risk of retinal artery occlusion and ischaemic stroke: a cross- cohort study*, by Wang et al., 2025, Br J Ophthalmol, 0, p. 5 (doi:10.1136/bjo-2023-325044). Copyright © 2025 by BMJ Group.

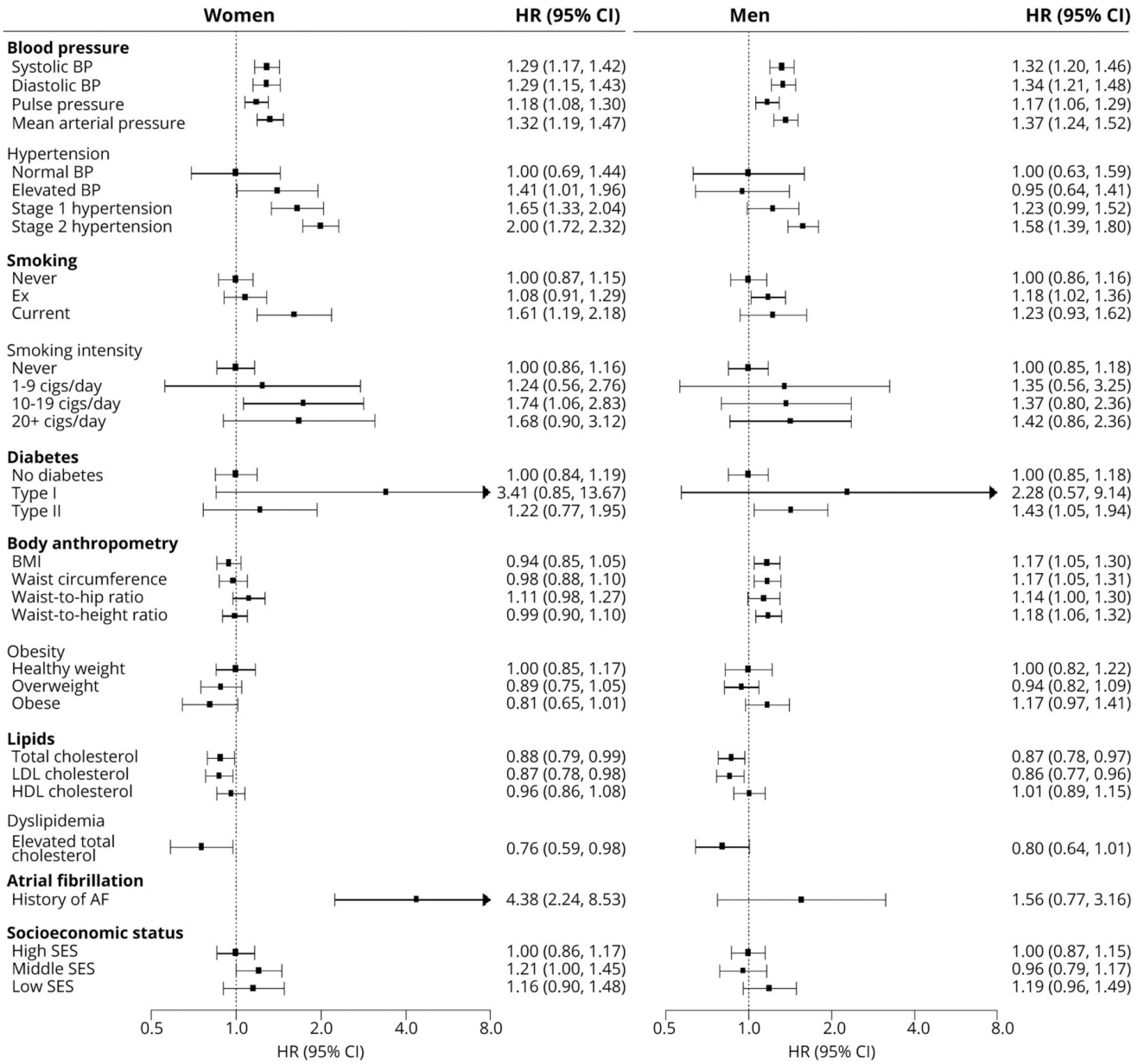
Figure C1: Adjusted HRs for incident ischaemic stroke associated with risk factors, by sex



Horizontal lines indicate the corresponding 95% confidence intervals (CIs) around the hazard ratios (HRs). AF = atrial fibrillation; BMI = body mass index; BP = blood pressure; HDL = high-density lipoprotein; LDL = low-density lipoprotein; SES = socioeconomic status.

Note: From *Sex differences in the association between major risk factors and the risk of stroke in the UK Biobank cohort study*, by Peters et al., 2020, Neurology, 95(20), p. e2719 (doi: 10.1212/WNL.0000000000010982). Copyright © 2020 by Wolters Kluwer Health, Inc. on behalf of the American Academy of Neurology.

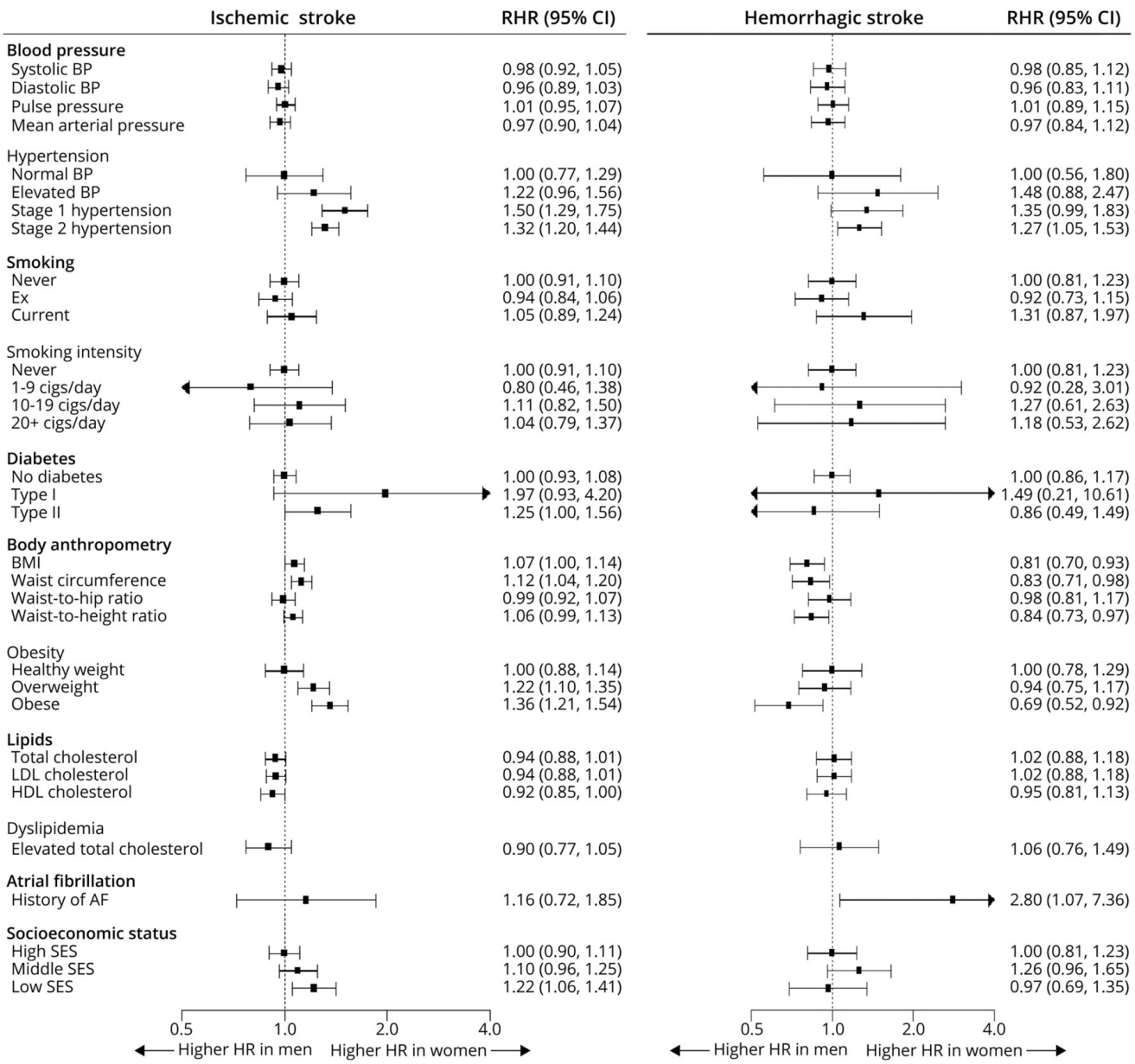
Figure C2: Adjusted HRs for incident haemorrhagic stroke associated with risk factors, by sex



Horizontal lines indicate the corresponding 95% confidence intervals (CIs) around the hazard ratios (HRs). AF = atrial fibrillation; BMI = body mass index; BP = blood pressure; HDL = high-density lipoprotein; LDL = low-density lipoprotein; SES = socioeconomic status.

Note: From *Sex differences in the association between major risk factors and the risk of stroke in the UK Biobank cohort study*, by Peters et al., 2020, Neurology, 95(20), p. e2720 (doi: 10.1212/WNL.0000000000010982). Copyright © 2020 by Wolters Kluwer Health, Inc. on behalf of the American Academy of Neurology.

Figure C3: Adjusted women-to-men RHRs for ischaemic and haemorrhagic stroke associated with risk factors



Horizontal lines indicate the corresponding 95% confidence intervals (CIs) around the ratio of hazard ratios (RHRs). AF = atrial fibrillation; BMI = body mass index; BP = blood pressure; HDL = high-density lipoprotein; LDL = low-density lipoprotein; SES = socioeconomic status.

Note: From *Sex differences in the association between major risk factors and the risk of stroke in the UK Biobank cohort study*, by Peters et al., 2020, Neurology, 95(20), p. e2719 (doi: 10.1212/WNL.0000000000010982). Copyright © 2021 by Wolters Kluwer Health, Inc. on behalf of the American Academy of Neurology.